Final Report
Investigation and tool development for storing NASA ECS data using HDF5 Archival Information

September 2009

Principal Investigator: MuQun (Kent) Yang
Institution: The HDF Group, 1901 So. First St., Suite C-2, Champaign IL 61820
CO-Investigator: Ruth Duerr at NSIDC
Period covered: July 2007–July 2009
NOAA Award Number: NA07OAR4310286

The original goals of this award are:

(1) To enhance the HDF4 to HDF5 conversion tool so that an HDF-EOS2 file can be converted to an HDF5 file that can be accessed by netCDF-4

(2) To design and implement an ECS granule level metadata to HDF5 METS conversion tool

(3) To design and implement an ECS collection level metadata to NMMR conversion

During the project, a number of changes have occurred. First, ISO-19115 has become the direction that NOAA is heading for metadata records. Second, the FGDC-based NMMR system is being replaced by a newer ISO-based system. Consequently, upon consultation with our colleagues at NGDC and on the NOAA Metadata ITAT group, we have decided to slightly modify the second and the third goals. The second and the third goals become:

(2) To design and implement an ECS granule metadata to METS conversion tool that is ISO-19115 and PREMIS compliant

(3) To design and implement an NSIDC/ECS data set metadata to ISO-19115 conversion tool

The following research summaries are organized to correspond with the revised goals.

(1) To enhance the HDF4 to HDF5 conversion tool

We have enhanced the HDF4 to HDF5 conversion tool so that HDF-EOS Grid and Swath files can be converted to HDF5 files that can be accessed by the netCDF-4 library. CF conventions are followed to ensure that the netCDF-4 library can read the geo-location information in the converter file. The tool has been released through The HDF Group website. We have also implemented a prototype verification tool that can be used to check if the HDF-EOS2 file is correctly converted. We have presented this work at the HDF and HDF-EOS workshop and the AMS annual meeting. Our experiences with retrieving the geo-location information from an HDF-EOS grid or swath will benefit future Common Data Model support for accessing NASA HDF-EOS data.

The conversion tool was tested not only against the HDF-EOS data products distributed by NSIDC but also against data products distributed by other NASA data centers such as GES DISC, LP-DAACs and LaRC. Several critical bugs were fixed and we are confident that the current release of the conversion tool can convert most NASA HDF-EOS2 swath and grid files successfully.

We also improved our algorithm to search the objects inside an HDF-EOS2 file. The overall performance of converting an HDF-EOS2 file is reasonable.

(2) To design and implement an ECS granule metadata to METS conversion tool that is ISO-19115 and PREMIS compliant

A tool that accesses MODIS data stored in NSIDC's ECS archive and transforms it into a complete HDF-Archive Information Package was completed. The tool generates valid METS, PREMIS, and ISO-19115 compliant XML metadata and transforms the MODIS HDF-EOS2 data into HDF-5 / netCDF 4 data. Because there are currently no publicly accessible mechanisms for discovering or accessing ancillary files such as production history or quality summary file, the tool doesn't support NSIDC's AMSR-E data product. NASA's ECS contractor, Raytheon, is preparing and update to the ECS system that will provide that accessibility. A potential future path forward would be to extend the tool to handle all ECS EOS-HDF 2/HDF4 data once these system changes are in place.

(3) To design and implement an NSIDC/ECS data set metadata to ISO-19115 conversion tool

A map between NSIDC's existing metadata catalog and the ISO-19115 metadata standard was generated and used to extend an existing web service to include the ability to output ISO-19115 XML, html, and text metadata for any data set in NSIDC's data catalog. The data catalog includes all of NSIDC's ECS data sets so extends the work beyond that proposed. The tool is currently undergoing testing by NSIDC's catalog team and the broader NSIDC community prior to installation into NSIDC's operational system. The metadata produced has been validated using the published ISO schema definitions. As might be expected some level of ambiguity was found during the mapping process and interpretation was needed to map the existing metadata into the new schema. Individual product managers reviewed the results for their own products and minor updates were requested. In addition, samples of the metadata have been provided to other NOAA colleagues such as Ted Habermann for review.

Appendix: Publications, Presentations, Posters and website

Presentations

"Developing Archive Information Packages for Data Sets: Early Experiments with Digital Library Standards", Eos Trans. AGU, 89(53), Fall Meet. Suppl., Abstract U13D-04, Ruth Duerr, MuQun Yang, M. Gooyabadi, Choonghwan Lee

"Creating Archive Information Packages for Data Sets: Early Experiments with Digital Library Standards", Ruth Duerr, MuQun Yang, A. Sikander and Choonghwan Lee, ESIP Federation meeting, July 8, 2009, Santa Barbara, California

"Using HDF5 Archive Information Package to preserve HDF-EOS2 data", Kent Yang and Ruth Duerr, The 12th HDF and HDF-EOS workshop, November, 2008, Denver, Colorado, http://hdfeos.org/workshops/ws12/presentations/day2/rxd2.ppt

"NOAA Science Data Stewardship project update", Kent Yang, The 11th HDF and HDF-EOS workshop, November, 2007, Landover, Maryland, http://hdfeos.org/workshops/ws11/presentations/day3/NOAA-SDS-Final.ppt

Publications:

"Investigation of Using HDF5 Archival Information Packages (AIP) to store NASA ECS data", MuQun Yang, Ruth Duerr and Choonghwan Lee, the 89th AMS annual meeting, January, 2009, Phoenix, Arizona

http://ams.confex.com/ams/pdfpapers/148062.pdf

"Towards a standard archival format for Earth science data: Storing NASA ECS data using HDF5 Archival Information Packages (AIP)" , Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, Ruth Duerr, MuQun Yang and Choonghwan Lee, 2008

Website:

Project website: http://www.hdfgroup.org/projects/noaa-sds

Enhanced HDF4 to HDF5 conversion tool: http://www.hdfgroup.org/h4toh5/

ECS granule to METS metadata conversion tool: Contact Ruth Duerr at NSIDC rduerr@nsidc.org

ECS data set to ISO-19115 metadata conversion tool: Contact Ruth Duerr at NSIDC rduerr@nsidc.org